By: Alham O. Hotaki

October 2022

US Census Demographic Data

Tableau Project

This data comes from a Kaggle dataset, it includes the census data for all counties in 2015.

It includes data from the entire country. The expanded data will allow for much more interesting analyses and will also be much more useful at supporting other data sets.

Counties are political subdivisions, and the boundaries of some have been set for centuries. Census tracts, however, are defined by the census bureau and will have a much more consistent size. A typical census tract has around 5000 or so residents.

The data here are taken from the DP03 and DP05 tables of the 2015 American Community Survey 5-year estimates. As a product of the US federal government, this is not subject to copyright within the US.

Questions and Insights to be analysed and prepared in
Tableau and presented via at least three visualisations.

01

Which states have the best transportation?

This is a subjective question, so your first job is to define what the best transportation is. Is it highest percentage of transit use? Is it lowest mean commute times. Then you need to determine how to aggregate the data from the county level to the state. Are there outlier counties affecting the data? How should you aggregate all the data from the counties to represent the state effectively? Please provide your reasoning in your report.

02

How does income and poverty look across America?

Think about how best to contrast this data to show an interesting finding. You can look across many of the different fields to show interesting findings. Do counties with more construction experience more or less poverty? Do counties near the coast experience more or less income? Remember this is all correlation and not causation so we cannot say any one thing causes it but we can report descriptive statistics.

03

Any other interesting topic to be analysed and presented.

As you work with the data, come up with a question you're curious about and can be answered from the data. Build a dashboard or story to answer your question and lead viewers to that answer.

Which states have the best transportation?

Insights:

Measure 1: Using public transport is a measure of sound infrastructure for public transportation. At an initial look to the data, the following is an overview of transportation status by states:
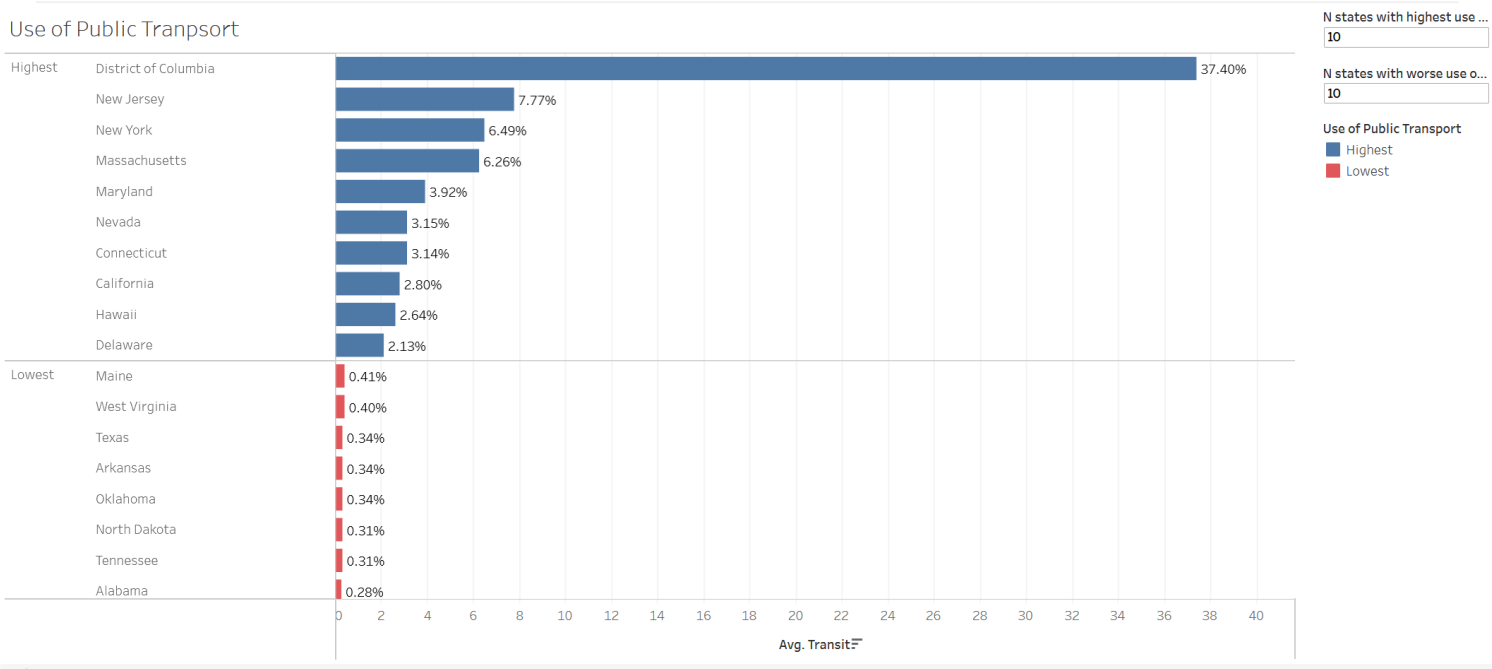

Figure 1 Use of Public Transport by States

As it is seen, top 10 states have a wide different range of average use of public transport. 37.4% of population as average use public transport in District of Columbia while this state seems like an outlier to the rest of states. At the second place, New Jersey's 7.77% of population as average uses public transportation. From bottom, Alabama is the bottom in the list where only 0.28% of population as average uses public transport.

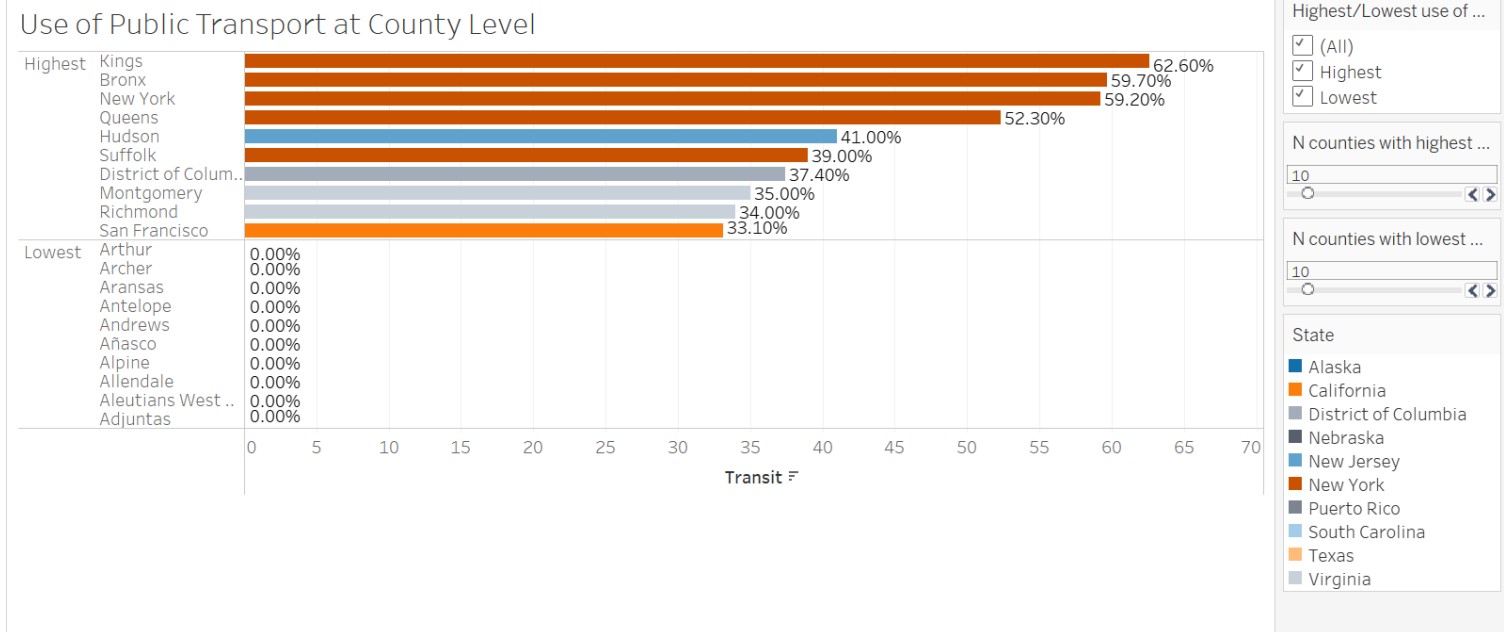Here is another look at the data, not at state level, but at County level:

Figure 2 Use of public transport by counties

It is well shown that highest use of public transport in counties are not distributed across all states. Four counties with highest use of public transport are in New York State, subsequently New Jersey, Virginia, and California, if we ignore DC as an outlier.

By excluding counties with highest use of public transport and considering them as outliers, the overall average of use of public transport across states will be changed. This will also change the ranking of states from transportation point of view as well.
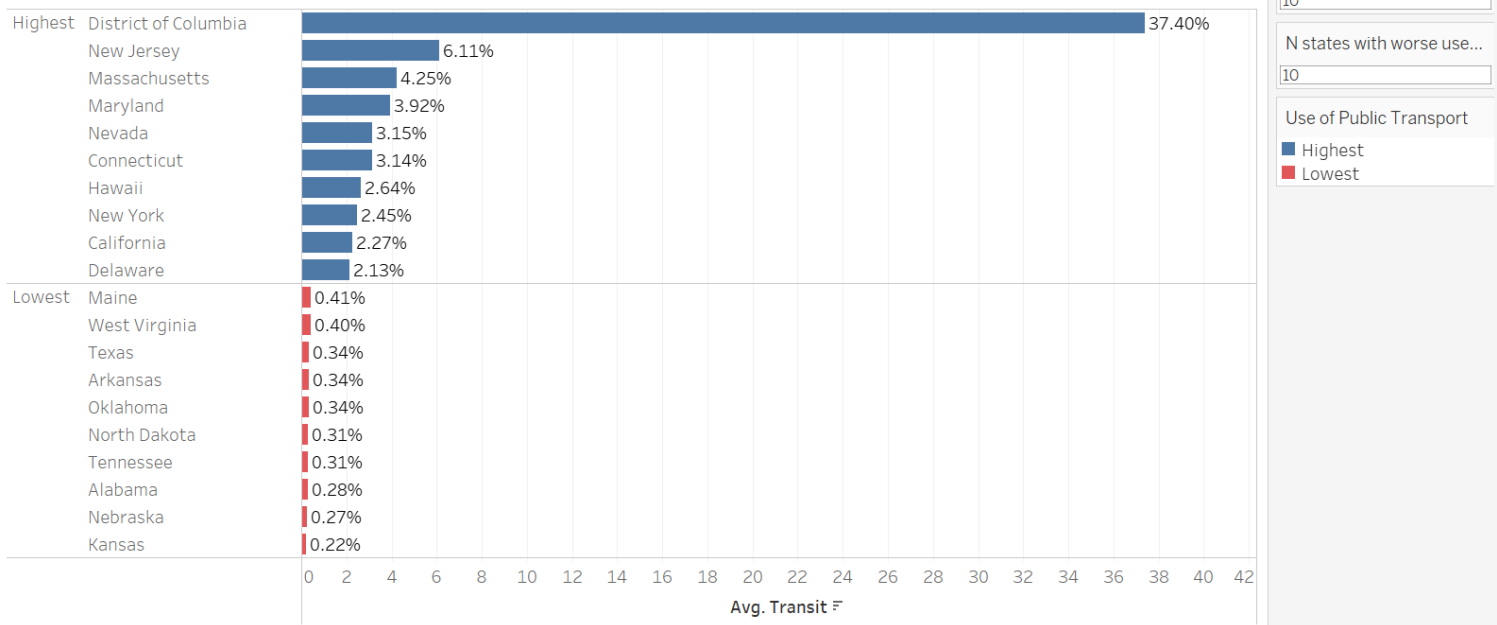

Figure 3 Use of Public Transport by State Level (without Outliers)

As it is seen above, New Jersey saves its place while New York state by excluding the outliers, downranked from third place to 8th place in the chart. Similarly, California downgrades from ranking 8 to ranking 9. Hawaii from place 9 upgrades to ranking 7.

Measure 2: Commute time is another indicator to measure the efficiency of public transport. The lower the commute time, the most efficient is the public transport. But this indicator alone can be evaluated and relied upon. This is also a subjective measure. For instance, there are states where use of public transport is less than 1% while commute time is less than the average. This alone cant be measured as efficient public transport indicator. To have a proper look at the public transport situation across the states, it would be more logical to have a look at the commute time indicator along with average use of public transport.
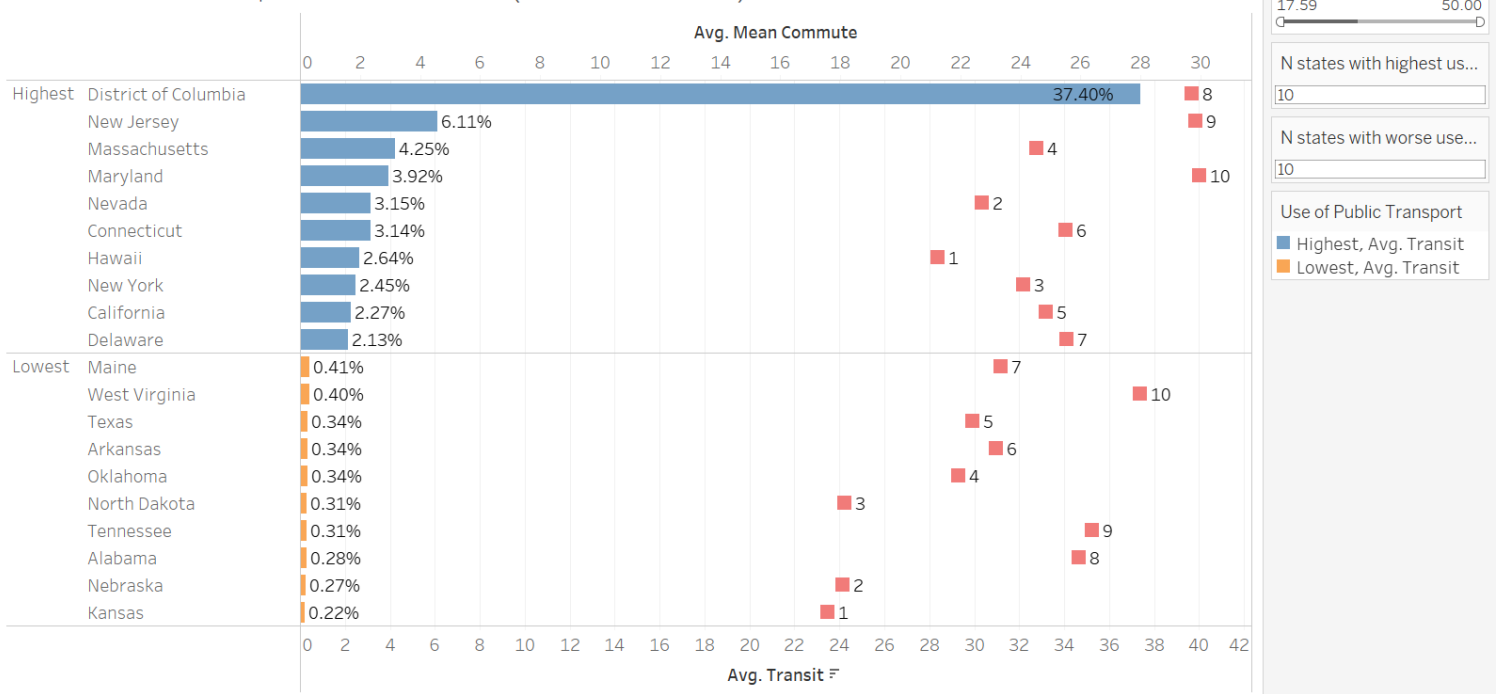
Figure 4 Use of Public Transport at state level and commute time

So, looking at top states (excluding outlier counties) from use of public transport point of view and then assessing based on the commute time ranking, the above picture gives a fairly indicative measure on the status of transportation in top ten states. Always to remember that DC is an overall outlier.

Design:

Creation of parameters and then creation of sets and linking each set with each parameter and for better filtering, creation of sets of sets were some of the methods used to create this story.

Colours: When choosing colours for the above charts, following principles were kept into consideration:

- Less intense colours
- To highlight the message behind the charts

Selection of reddish colours for negative insights (such as states or counties with less use of public transport) or highlighting a variable, and then tried to not use green with red, instead blue (calmer palette) was used with red which is more colour-blind friendly palette. In chart on counties, the whole palette was selected as colour-blind friendly.

Bar charts: selection of this type of chart was only because a categorical variable was comparing with a quantitative variable. As well as if we are comparing more than two variables, then bar charts are ideal. This may be not the only right solution, but it was preferred in this project.

How does income and poverty look across America?

Insights:

Income and poverty are correlated. The highest Income, the lowest the poverty. In order to have an overview of poverty across all the states, it would be better to have a look to the following map:
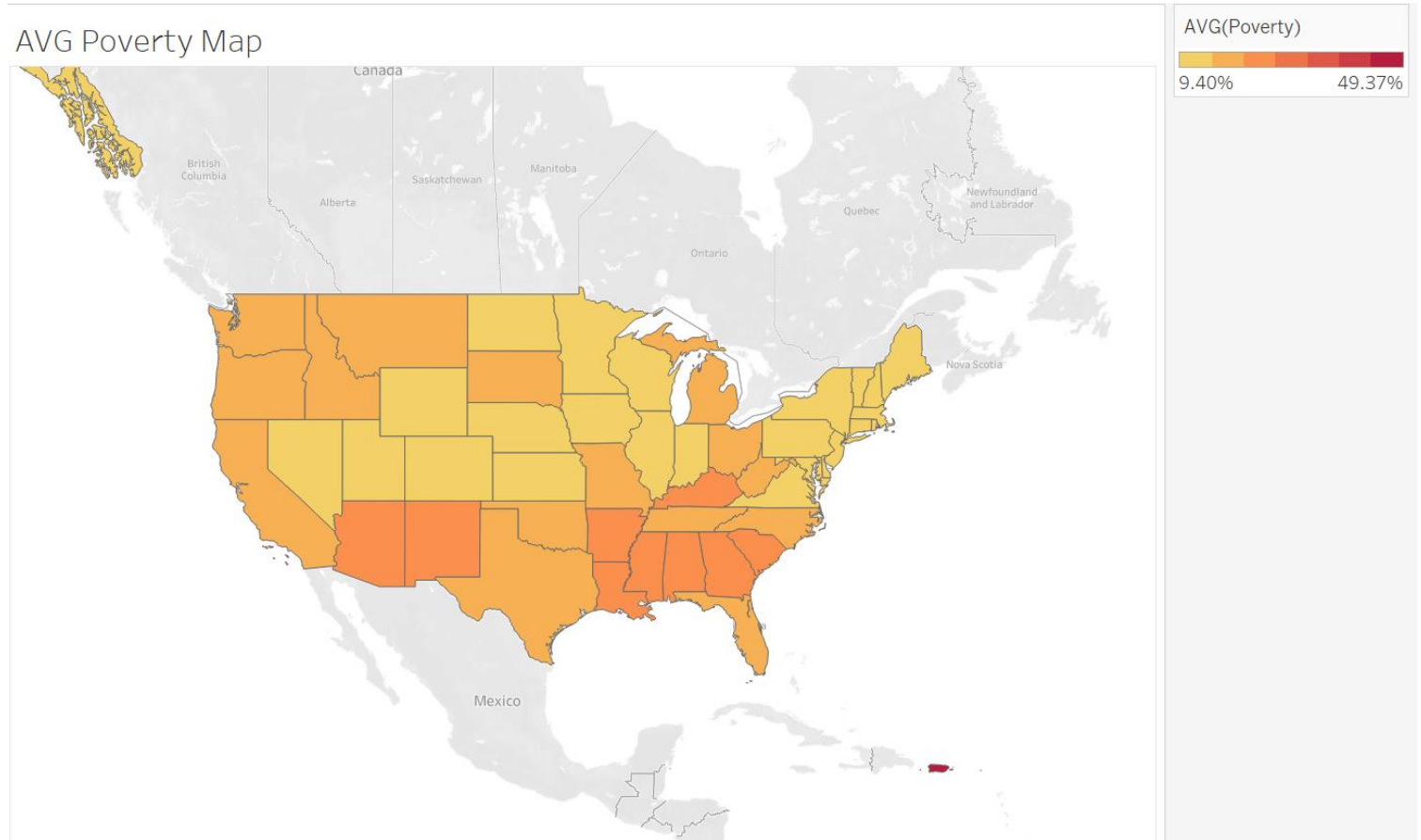

Figure 5 Average Poverty Map

It shows that there is no state with less than 15% average poverty in southern region. It also shows that Puerto Rico is the poorest among all other states and acts as an outlier in this analysis.

The following chart shows that least poor and most rich states:

N Rich/Poor states

State	Avg. Poverty
Connecticut	9.4%
New Hampshire	10.2%
New Jersey	10.5%
Rhode Island	10.7%
Maryland	11.2%
North Dakota	11.4%
Minnesota	11.7%
Wyoming	11.7%
Vermont	11.8%
Iowa	11.9%
New Mexico	21.1%
South Carolina	21.2%
Arizona	21.3%
Arkansas	21.6%
Alabama	21.8%
Kentucky	21.8%
Louisiana	22.0%
Georgia	22.2%
Mississippi	25.7%
Puerto Rico	49.4%

N poor States
10

N rich states
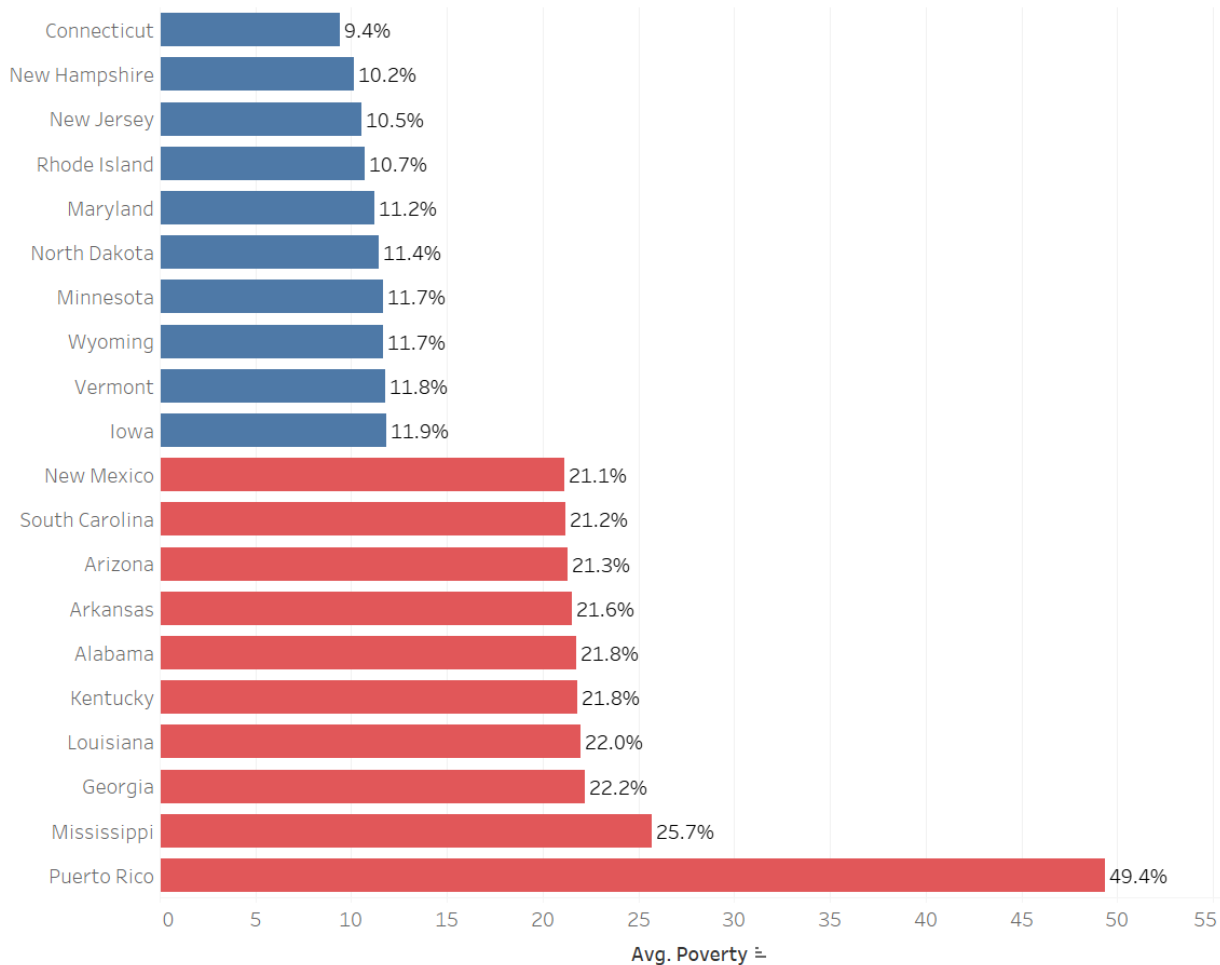10

Poverty accross states
■ Rich
■ Poor

Figure 6 N Rich/Poor States

Looking at income and poverty scatter plot, it clearly defines that there is a strong negative correlation existing. This also can be shown at county level and besides that, it could be very useful if we cluster all counties into four categories.

Clusters 1 and 3 would be Mid – High and Mid-Low, cluster 4 – the poorest and cluster 2 – the richest states from income/poverty point of view. The following plot defines afore-mentioned statement very well:

Poverty/Income Clustering
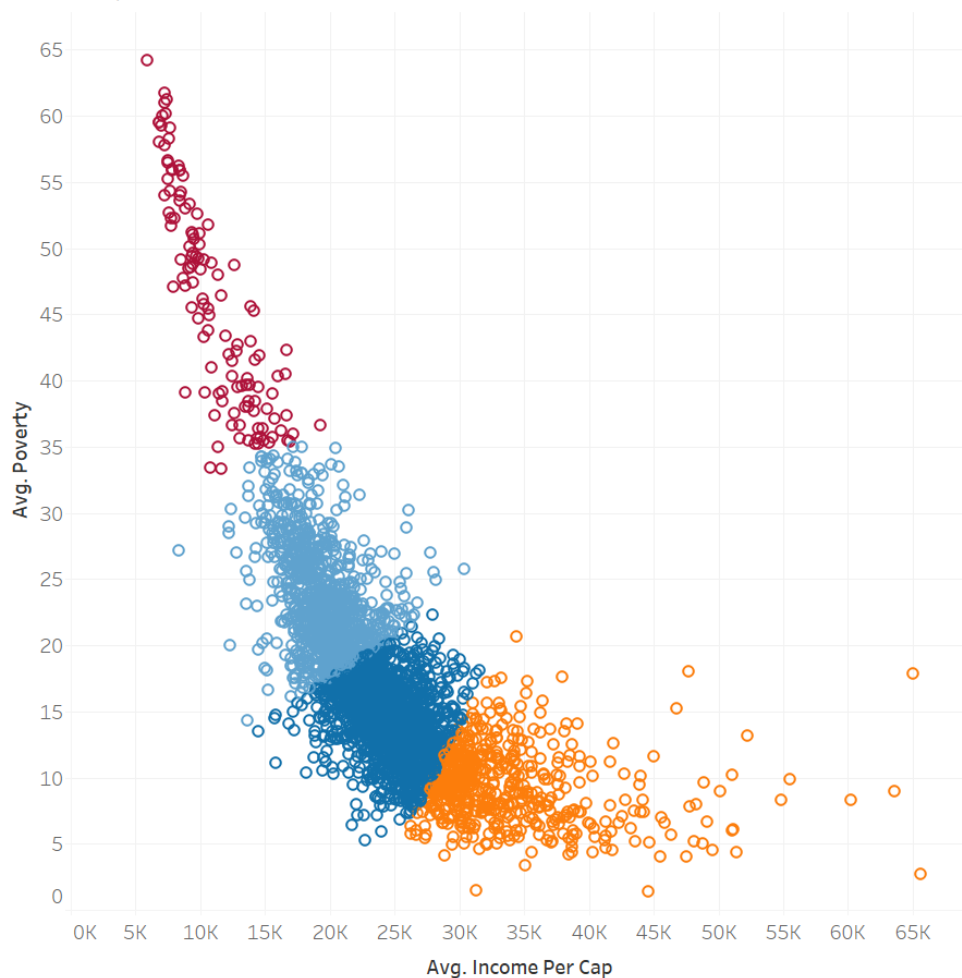
Figure 7 Poverty/Income clustering

Is there a correlation between poverty and different ethnicities? In order to get this insight, following scatter plots are worth to have a look:
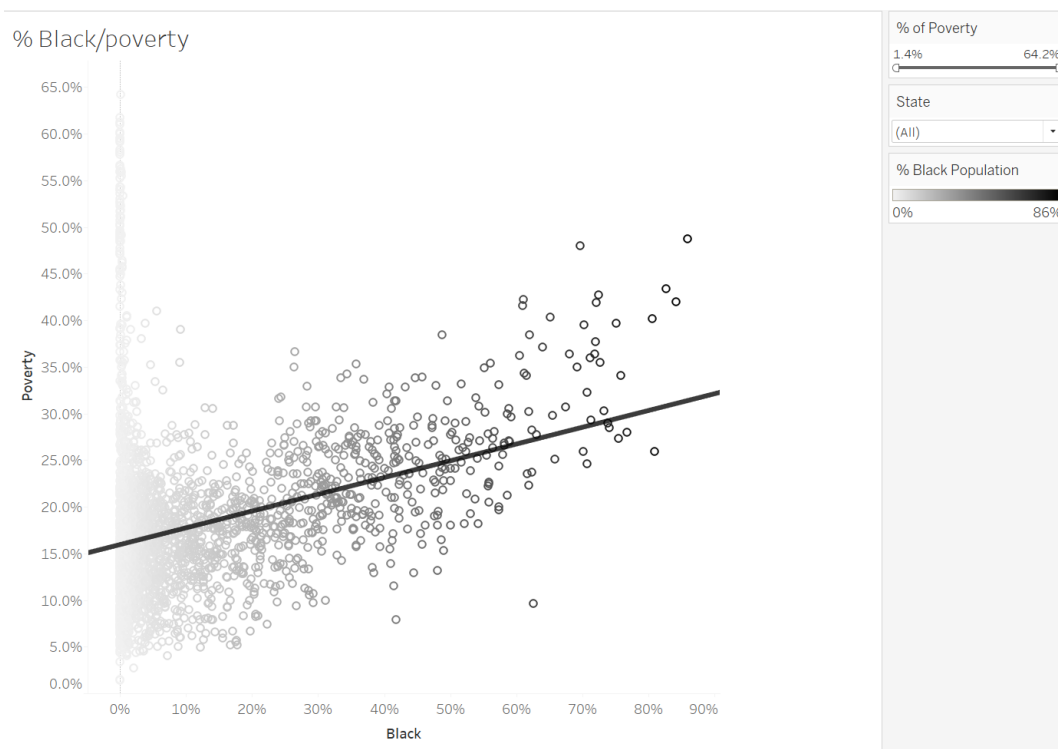

Figure 8 % Black Population/Poverty

It seems that there is a strong positive correlation between percentage of poverty and percentage of black population across counties.
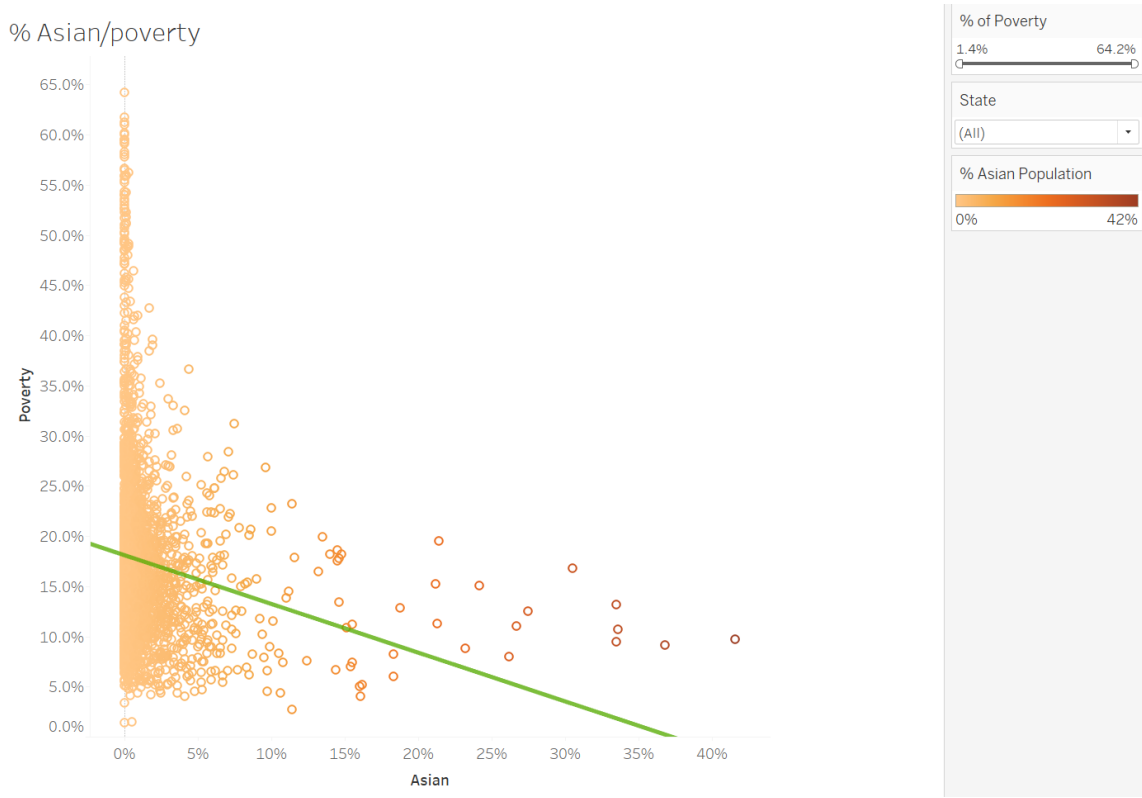

Figure 9 %Asian Population/Poverty

It seems that there is a negative correlation between percentage of poverty and percentage of Asian population across counties. Means, more Asian population are less under poverty line.
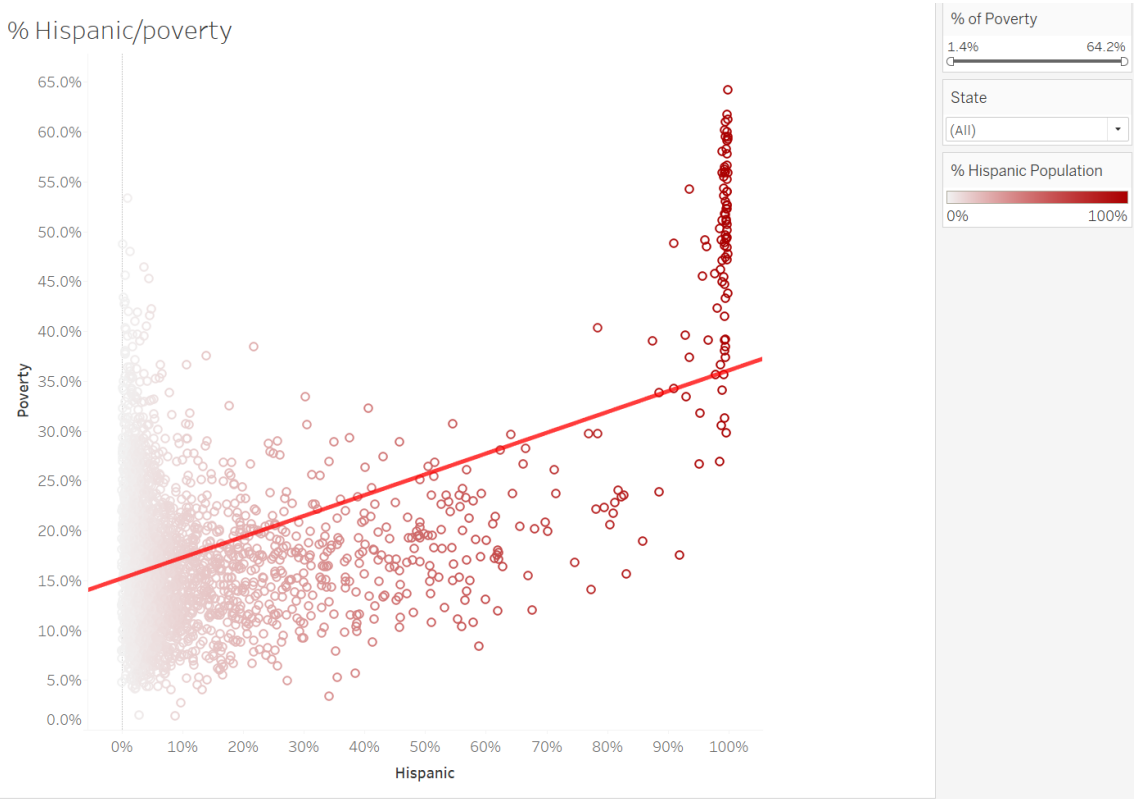

Figure 10 % Hispanic Population/Poverty

It seems that there is a strong positive correlation between percentage of poverty and percentage of Hispanic population across counties. Means, more Hispanic population are more under poverty line.
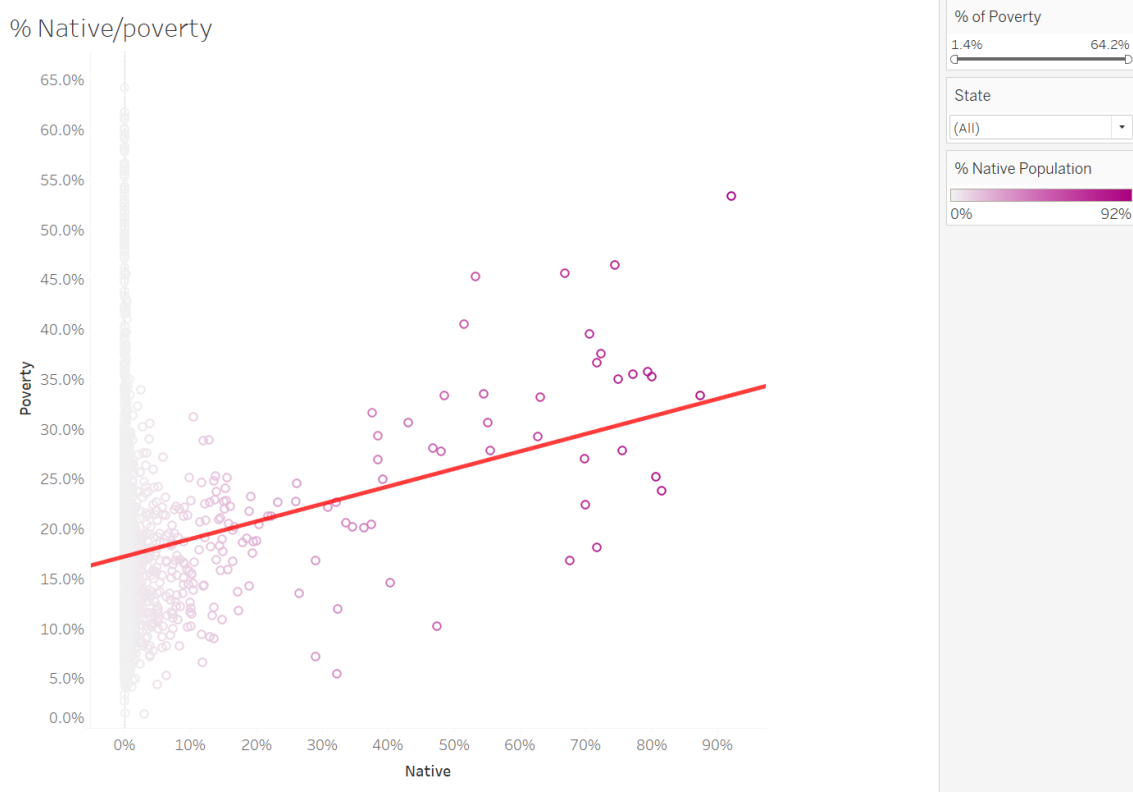
Figure 11 %Native Population/Poverty

It seems that there is a positive correlation between percentage of poverty and percentage of Native population across counties. Means, more Native population are more under poverty line.
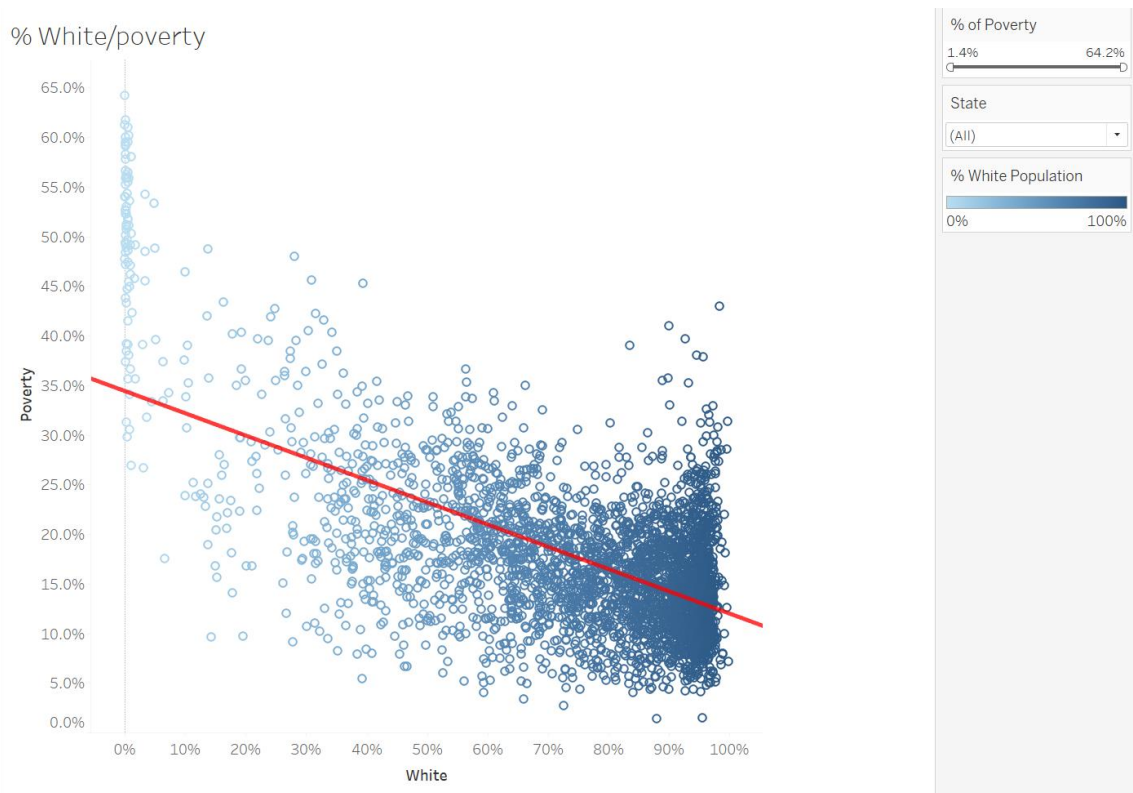

Figure 12 %white population/poverty

It seems that there is a strong negative correlation between percentage of poverty and percentage of white population across counties. Means, more white population are less under poverty line.

Design:

Creation of parameters, sets and sets of sets from the previous section has been continued. While creation of map and using it as a filter, also creation of scatter plots with trend line and using clusters were other aspects used in creation of this story.

Charts: selection of map chart was to provide an overview of all the states at once. Selection of bar chart was to compare categorical variable with quantitative value and more than two variables. However, it was not the only solution, but it was preferred during this project. Selection of scatter plots were to provide two insights: strength and direction of correlation. While taking into consideration that correlation is not the causation, scatter plots provide better overviews to compare two quantitative values. Clustering was another chart produced based on the scatter plot in this section. The clustering was to provide just an overview on layers of population if divided by poverty and income level.

Colours: the colours were selected in a way to tell the background insight as well, for example, reddish or darker – defining negative aspects, lighter palettes – defining the positive story behind the charts. For map chart – red/gold palette was selected to be more colour-blind friendly, lighter colours – rich states, darker colours – poor states. Red colour was used in bar chart to convey the negative aspect of poor states while was not used with greenish palettes to avoid the colour – blindness problem. Colour selection for Scatter plot clustering was based on the colour-blind friendly principle – choosing the palettes between blue and orange adding red to the most negative aspect of the insight (most poor counties). For other three scatter plots (ethnicities), it was attempted to provide more colour-blind friendly colours while considering the nature of the ethnicity as well.

Nature of Jobs

Links:

Insights:

Let's have a look, which states are employed in which sectors. For this we will identify the highest percentage of the nature of the job in that state at an average and then have a look at the income per capita for that state as an average. This will give us a subjective insight on which jobs are well paid and which jobs are playing key role in higher income for that specific state. We will exclude two regions as outliers: DC and Puerto Rico.

State	Avg. Income Per Cap	% of employed in construction (Ranking)	% of employed in office works (Ranking)	% of employed in private work (Ranking)	% of employed in Production (Ranking)	% of employed in Professional work (Ranking)	% of employed in public works (Ranking)	% of self employed (Ranking)	% of employed in service works (Ranking)	% of employed in work at home (Ranking)
Rhode Island	35,671	8	3	2	8	1	7	7	4	6
North Dakota	31,901	2	8	5	3	3	6	1	8	2
Hawaii	31,893	7	2	7	7	5	2	3	1	4
Alaska	30,333	3	7	8	2	4	1	4	5	3
Nevada	26,318	1	5	4	5	7	3	5	3	7
South Dakota	25,335	4	6	6	4	2	4	2	6	1
Indiana	23,977	6	4	1	1	8	8	8	7	8
Florida	23,502	5	1	3	6	6	5	6	2	5

Figure 13 Nature of jobs and associated level of income per capita

As it is seen, Rhode Island as the top from average income per capita point of view (excluding outliers), Ranking number 1 from employment point of view is Professional work (management, business, science, and arts). Florida is the last in the table having ranking number 1 in office works. Looking at income per capita from the highest to lowest and matching it with the ranks of each type of works, it can be subjectively showing the categories of payments under each job.

The following chart shows that in which type of jobs, most of the population are employed:
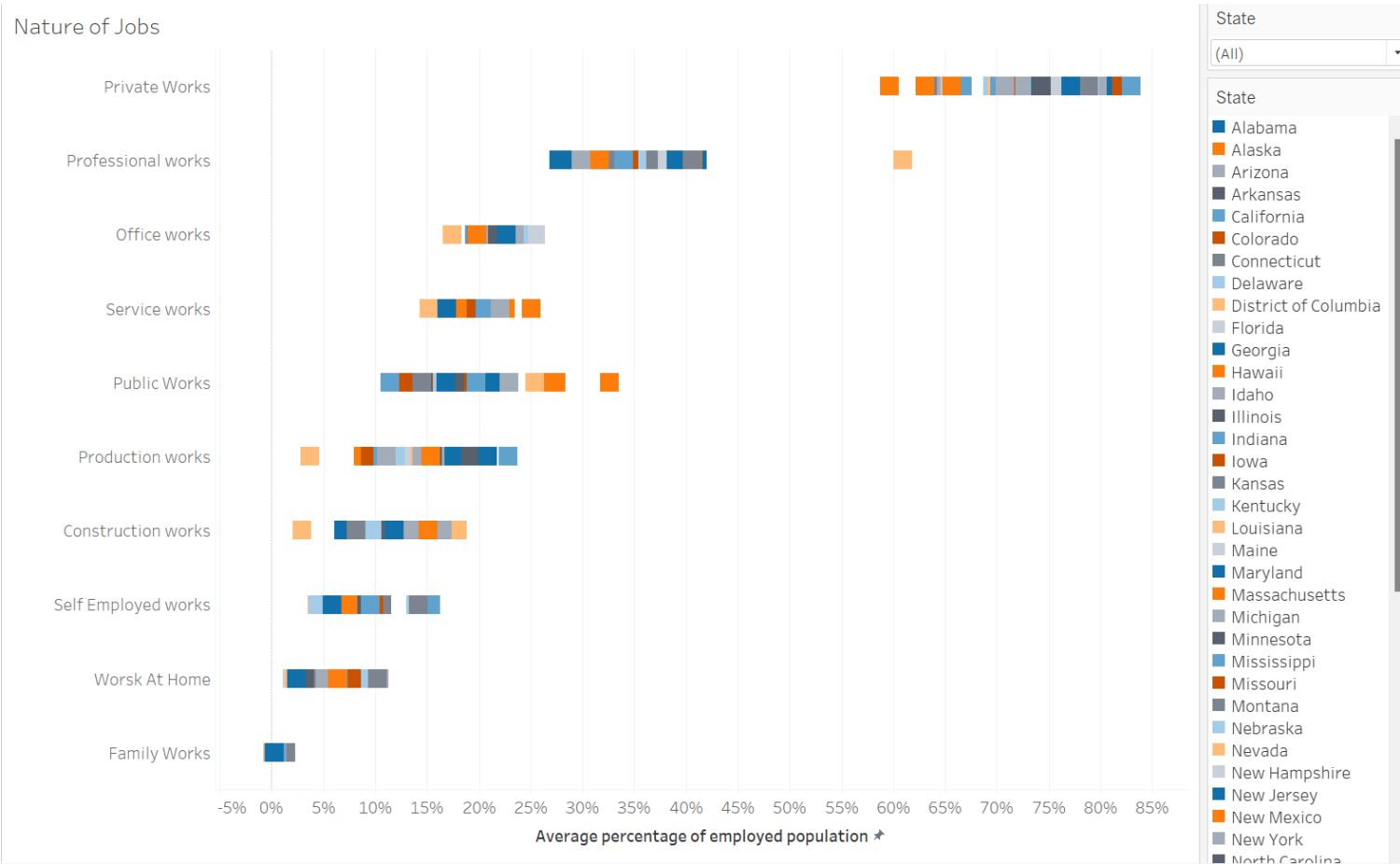

Figure 14 highest to lowest percentage of population under each job category

The highest percentage of employed population are working in private companies while the lowest number of populations are working under Family works category which is unpaid work. Second highest category is the Professional sector including employment in management, business, science, and arts sectors. In this sector, DC again acts as an outlier, where an average 60% of DC employed population are working in this category.

Design:

For this story, one main table was selected intentionally. Sometimes, to attract more attention to the topic, it would be great to provide a challenge for the viewer to dig deeper into the slide provided to them. Making a small puzzle and when the viewer solves it by understanding the message behind it, it provides a feeling of self-satisfaction to the viewer and make the viewer more involved in the topic rather than just only being a viewer. The second bar chart was the only possible solution to compare three variables with two categorical components. The colour palette for the bar chart was selected colour blind friendly palette. However more colours seem similar, but the legend and highlighting action in legend will help to get insight from the chart.